

## Consistent Cytotoxic-T-Lymphocyte Targeting of Immunodominant Regions in Human Immunodeficiency Virus across Multiple Ethnicities

Nicole Frahm,<sup>1</sup> B. T. Korber,<sup>2,3</sup> C. M. Adams,<sup>1</sup> J. J. Szinger,<sup>2</sup> R. Draenert,<sup>1</sup> M. M. Addo,<sup>1</sup> M. E. Feeney,<sup>1</sup> K. Yusim,<sup>2</sup> K. Sango,<sup>1</sup> N. V. Brown,<sup>1</sup> D. SenGupta,<sup>1</sup> A. Piechocka-Trocha,<sup>1</sup> T. Simonis,<sup>4</sup> F. M. Marincola,<sup>4</sup> A. G. Wurcel,<sup>1,5</sup> D. R. Stone,<sup>5</sup> C. J. Russell,<sup>6</sup> P. Adolf,<sup>6</sup> D. Cohen,<sup>6</sup> T. Roach,<sup>7</sup> A. StJohn,<sup>7</sup> A. Khatri,<sup>8</sup> K. Davis,<sup>9</sup> J. Mullins,<sup>9</sup> P. J. R. Goulder,<sup>1</sup> B. D. Walker,<sup>1</sup> and C. Brander<sup>1\*</sup>

*Partners AIDS Research Center<sup>1</sup> and Endocrine Unit,<sup>8</sup> Massachusetts General Hospital, Charlestown, and Fenway Community Health Center<sup>6</sup> and Lemuel Shattuck Hospital, Boston,<sup>5</sup> Massachusetts; Clinical Center, National Institutes of Health, Bethesda, Maryland<sup>4</sup>; Queen Elizabeth Hospital, Bridgetown, Barbados<sup>7</sup>; Los Alamos National Laboratory, Los Alamos,<sup>2</sup> and Santa Fe Institute, Santa Fe,<sup>3</sup> New Mexico; and Department of Microbiology, University of Washington, Seattle, Washington<sup>9</sup>*

Received 31 July 2003/Accepted 13 November 2003

Although there is increasing evidence that virus-specific cytotoxic-T-lymphocyte (CTL) responses play an important role in the control of human immunodeficiency virus (HIV) replication *in vivo*, only scarce CTL data are available for the ethnic populations currently most affected by the epidemic. In this study, we examined the CD8<sup>+</sup>-T-cell responses in African-American, Caucasian, Hispanic, and Caribbean populations in which clade B virus dominates and analyzed the potential factors influencing immune recognition. Total HIV-specific CD8<sup>+</sup>-T-cell responses were determined by enzyme-linked immunospot assays in 150 HIV-infected individuals by using a clade B consensus sequence peptide set spanning all HIV proteins. A total of 88% of the 410 tested peptides were recognized, and Nef- and Gag-specific responses dominated the total response for each ethnicity in terms of both breadth and magnitude. Three dominantly targeted regions within these proteins that were recognized by >90% of individuals in each ethnicity were identified. Overall, the total breadth and magnitude of CD8<sup>+</sup>-T-cell responses correlated with individuals' CD4 counts but not with viral loads. The frequency of recognition for each peptide was highly correlated with the relative conservation of the peptide sequence, the presence of predicted immunoproteasomal cleavage sites within the C-terminal half of the peptide, and a reduced frequency of amino acids that impair binding of optimal epitopes to the restricting class I molecules. The present study thus identifies factors that contribute to the immunogenicity of these highly targeted and relatively conserved sequences in HIV that may represent promising vaccine candidates for ethnically heterogeneous populations.

Cytotoxic T lymphocytes (CTL) recognizing HLA class I-restricted viral epitopes are considered an important arm of the antiviral immune defense in human immunodeficiency virus (HIV) infection (11). This is supported by CTL depletion and immune-escape studies, as well as vaccination experiments in the simian immunodeficiency virus model, which suggest that robust CTL activity can potentially control infection (5, 6, 49). In humans, the temporal association of CTL responses with reduction in the viral load in early HIV infection and the association of HIV disease progression with certain HLA class I alleles and supertypes reinforce the findings of animal studies (33, 37, 43, 51, 52; E. Trachtenberg, B. Korber, C. Sollars, E. Hayes, R. Funkhouser, T. Kepler, P. Hraber, M. Hsu, H. Erlich, and S. Wolinsky, abstract from the ASHI 28th Annu. Meet., Hum. Immunol. 63:S32, 2002). However, even broad CTL activity may not be able to avert subsequent superinfection (3), and early and late escape from CTL responses can be associated with increases in the viral load (27, 35). Such data

suggest that CTL responses to relatively conserved regions will be critical for the development of an effective vaccine, even though the inclusion of variable regions may also provide some benefit (44).

The identification of highly immunogenic and conserved regions is complicated by HIV sequence diversity and differences in host genetics among different ethnic groups (17). Even in geographically restricted regions, such as large urban areas, where one HIV clade dominates the epidemic, the ethnic distribution and HLA allele variability may complicate the selection of a promising vaccine candidate, particularly when subunit- or epitope-based vaccines are being considered. In addition, many studies, especially those performed early in the HIV epidemic in the United States, have focused on HIV-infected individuals of Caucasian descent, leading to a paucity of CTL data for the ethnicities in which the HIV epidemic currently spreads fastest, including African, Asian, and Caribbean populations (1, 7, 16, 23, 31, 41, 42, 46).

The present study was designed to identify regions in HIV clade B protein sequences that are frequently targeted by HIV-infected individuals from different ethnicities and which may therefore be particularly relevant for clade B-based vaccine de-

\* Corresponding author. Mailing address: Partners AIDS Research Center, 5th Floor MGH East, No. 5239, 149 13th St., Charlestown, MA 02129-2000. Phone: (617) 724-5789. Fax: (617) 726-5411. E-mail: brander@helix.mgh.harvard.edu.

TABLE 1. Demographics, viral loads, and CD4 counts of 150 study subjects

Parameter	Value <sup>a</sup>			
	African-Americans	Caucasians	Hispanics	West Indians
<i>n</i>	59	26	44	21
Gender (male/female)	45/14	20/6	32/12	11/10
Treatment (on/off)	43/16	15/11	24/19	6/15
Median viral load	497	272	5,654	11,300
No. of RNA copies/ml	(<50–400,000)	(<50–106,000)	(<50–>750,000)	(<50–>750,000)
Median CD4 count	329	322	283	223
No. of cells/ $\mu$ l	(7–1,280)	(28–945)	(11–1,476)	(3–792)

<sup>a</sup> Numbers in parentheses indicate ranges of values observed among enrolled individuals.

sign and testing. Initially, the HLA allele distributions among HIV-infected and uninfected African-American, Hispanic, or West Indian individuals living in the Boston area and in HIV-infected individuals in Barbados were determined and compared to those in Caucasians by performing intermediate- to high-resolution HLA typing. One hundred fifty HIV-infected individuals were then comprehensively tested for total HIV-specific CTL activity using a clade B consensus sequence-based overlapping-peptide (OLP) set in enzyme-linked immunospot (ELISpot) and intracellular cytokine-staining analyses. The most immunogenic regions across all tested ethnicities were then determined, and factors that might contribute to the immunodominance of those regions were investigated.

#### MATERIALS AND METHODS

**Study subjects.** A total of 275 HIV-positive and HIV-negative subjects were recruited for the HLA-typing study from four hospitals in the Boston area and at the Queen Elizabeth Hospital in Barbados. Of these, 150 HIV-positive individuals at all clinical stages with detectable HIV-specific cellular responses were included in the CD8<sup>+</sup>-T-cell study; all persons were included regardless of CD4 count, viral load, or treatment status; 15 HIV-negative subjects were included as controls. The studies were approved by the institutional review boards of all participating hospitals, and all subjects provided written informed consent before being recruited. The study subjects' characteristics are summarized in Table 1.

**Lymphocyte separation.** Peripheral blood mononuclear cells (PBMCs) were separated from whole blood by density gradient centrifugation (Histopaque 1077; Sigma, St. Louis, Mo.) within 24 h of venipuncture for individuals recruited in the Boston area and were used directly in the *in vitro* assays described below. Blood samples from individuals recruited in Barbados were processed after transport to Boston, leading to a delay of 24 to 48 h in lymphocyte separation for 15 of the 21 samples. Of these, six were analyzed after being frozen, while for all other samples, the initial ELISpot screening and reconfirmation were carried out on freshly isolated PBMCs. Frozen samples were thawed, and the cells were rested in RPMI 1640 containing 10% heat-inactivated fetal calf serum (both from Sigma), 2 mM L-glutamine, 50 U of penicillin/ml, 50  $\mu$ g of streptomycin/ml, and 10 mM HEPES (all from Mediatech, Herndon, Va.) at 37°C and 5% CO<sub>2</sub> for 6 to 20 h prior to use in ELISpot assays.

**HLA typing.** High- and intermediate-resolution HLA class I typing was performed at the HLA-typing laboratory of the Department of Transfusion Medicine, National Cancer Institute, National Institutes of Health, by sequence-specific primer PCR as previously described (12). DNA for typing was extracted using the Puregene DNA isolation kit for blood (Gentra Systems, Minneapolis, Minn.) according to the manufacturer's instructions.

**HMA.** The heteroduplex mobility assay (HMA) was performed as previously described (20; <http://ubik.microbiol.washington.edu/HMA/index.html>). Briefly, PCR-amplified fragments corresponding to the C2-V5 region of HIV type 1 (HIV-1) Env from both the unknown specimen and a HIV-1 subtype reference were mixed, melted at 94°C, reannealed on ice, and then displayed by electrophoresis on a 5% neutral polyacrylamide gel, followed by staining with ethidium bromide. Heteroduplexes formed between the unknown sample and the most closely related reference sequences exhibit the fastest mobilities and thus identify the subtype of the unknown strain.

**Peptides.** A set of 410 OLPs was used to screen for HIV-specific T-cell responses. The peptides spanned all HIV proteins and were based on the con-

sensus B sequence of 2001 available at the HIV immunology database ([http://hiv-web.lanl.gov/content/hiv-db/CONSENSUS/M\\_GROUP/Consensus.html](http://hiv-web.lanl.gov/content/hiv-db/CONSENSUS/M_GROUP/Consensus.html)). The peptides were generally 18-mers, varying from 15 to 20 amino acids in length and overlapping by 10 amino acids. The length variation was caused by the exclusion of amino acids that are rarely found in C-terminal anchor motifs of HLA class I molecules from the C terminus (15). A previously described algorithm was applied for peptide design (<http://hiv-web.lanl.gov/content/hiv-db/PEPTGEN/2001.html>). This approach has recently been shown to provide a higher rate of peptide recognition than using an unadapted set of 18-mers (21). Peptides were synthesized at the peptide synthesis facility of the Massachusetts General Hospital using 9-fluorenylmethoxycarbonyl chemistry or were obtained from Research Genetics, Huntsville, Ala. The peptides were tested in a peptide pool matrix to facilitate comprehensive screening as described previously (1). Reconfirmations of all positive wells in the matrix screen were done the following day on a single-peptide basis. The numbering of peptides was as follows: Gag-p17 OLP 1 to OLP 17, Gag-p24 OLP 18 to OLP 48, Gag-p15 OLP 49 to OLP 66, Nef OLP 67 to OLP 93, Rev OLP 94 to OLP 108, Tat OLP 111 to OLP 122, Vpu OLP 123 to OLP 131, Pol-protease OLP 145 to OLP 164, Pol-reverse transcriptase (RT) OLP 165 to OLP 240, Pol-integrase OLP 241 to OLP 277, Vpr OLP 278 to OLP 288, Env-gp120 OLP 289 to OLP 365, Env-gp41 OLP 366 to OLP 401, and Vif OLP 402 to OLP 425.

**ELISpot assays.** PBMCs were plated in 96-well polyvinylidene difluoride plates (Millipore, Bedford, Mass.) which had been precoated overnight with 0.5  $\mu$ g of the anti-gamma interferon (IFN- $\gamma$ ) monoclonal antibody 1-D1K (Mabtech, Stockholm, Sweden)/ml. Each peptide was added at a final concentration of 14  $\mu$ g/ml (for both single peptides and pooled peptides). A total of 100,000 cells per well were added in 140  $\mu$ l of RPMI 1640 containing 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 50 U of penicillin/ml, 50  $\mu$ g of streptomycin/ml, and 10 mM HEPES. For negative controls, cells were incubated in medium alone. Phytohemagglutinin (Sigma) was added at a concentration of 1.8  $\mu$ g/ml to serve as a positive control. The cells were incubated overnight at 37°C with 5% CO<sub>2</sub>. The plates were developed by washing them six times with phosphate-buffered saline (no Ca or Mg; Mediatech), and 0.5  $\mu$ g of biotinylated anti-IFN- $\gamma$  monoclonal antibody 7-B6-1 (Mabtech)/ml was added for 1 h at room temperature. The plates were again washed and incubated with a 1:2,000 dilution of streptavidin-coupled alkaline phosphatase (Mabtech) for 1 h at room temperature in the dark. The plates were washed again, and IFN- $\gamma$  production was detected as blue spots after a short incubation with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad, Hercules, Calif.). The color reaction was stopped by washing the plates with tap water. The spots were counted using the AID ELISpot Reader unit (Autoimmun Diagnostika GmbH, Strasbourg, Germany), and the results were expressed as spot-forming cells (SFC) per million input cells.

Thresholds for positive responses were determined as at least five spots (50 SFC/10<sup>6</sup> PBMCs) per well and as responses exceeding the mean of negative wells plus 3 standard deviations and three times the mean, whichever was higher. In all tested subjects except three individuals, the mean of six or more negative control wells was <35 SFC/10<sup>6</sup> PBMCs, with an overall average of negative wells (including the high-background subjects) of 11 SFC/10<sup>6</sup> PBMCs. In addition to the above cutoff, responses were considered positive only if they were detected again in the subsequent confirmational ELISpot assay using single peptides. Two HIV<sup>+</sup> and all 15 of the HIV<sup>-</sup> individuals tested in a blinded fashion did not show detectable responses in ELISpot assays. Responses to peptides were previously shown to be mediated by CD8<sup>+</sup> T cells in depletion experiments using RosetteSep CD4 and CD8 reagents (StemCell Technologies, Vancouver, Canada) according to the manufacturer's directions (1, 23). Additionally, CD8<sup>+</sup>-T-cell dependence testing was performed on a subset of study subjects using flow

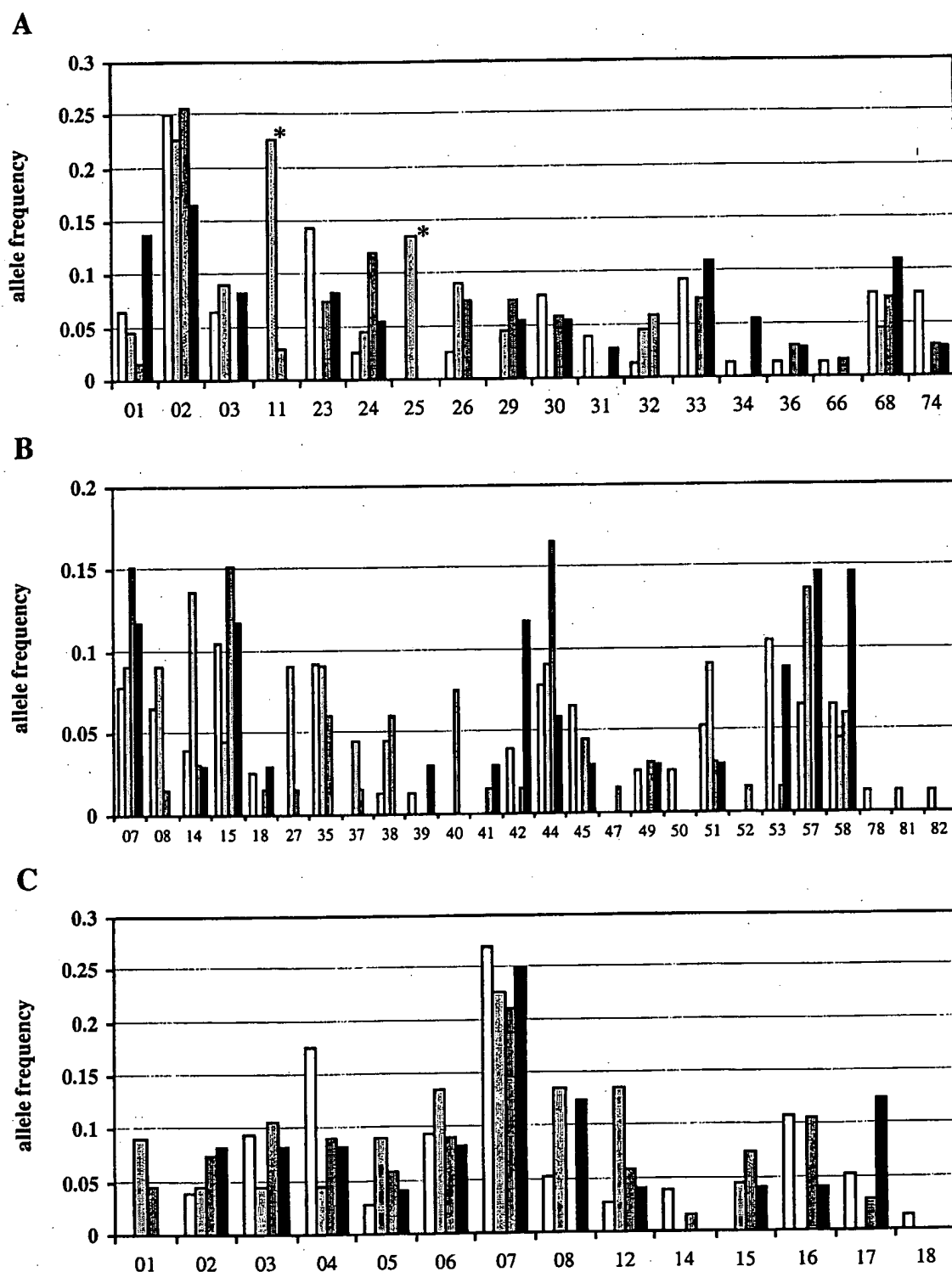


FIG. 1. HLA allele frequencies in different ethnicities. HLA class I allele distributions are shown for HLA-A (A), HLA-B (B), and HLA-Cw (C) alleles. Frequencies are shown for 73 African-Americans (open bars), 75 Caucasians (lightly shaded bars), 47 Hispanics (darkly shaded bars), and 80 West Indians (solid bars). The asterisks mark alleles that show significant differences among ethnicities (Fisher's exact test;  $P < 0.001$ ).

cytometric intracellular cytokine staining as previously described (8, 28) with the monoclonal antibodies anti-CD8-allophycocyanin, CD4-R-phycoerythrin, and IFN- $\gamma$ -fluorescein isothiocyanate (all from Becton Dickinson, San Jose, Calif.).

**Breadth and magnitude of responses.** The breadth of the responses was assessed both by considering every recognized peptide as a separate response

and by considering recognition of two adjacent peptides as a response to one "epitopic region" (1). To enumerate epitopic regions, the following algorithm was employed. When two adjacent peptides were recognized, only the stronger of the two responses was included. In the case of three adjacent peptides eliciting a response, the weakest of all three peptides was deleted, and the responses were

TABLE 2. Distribution of CTL responses within and between proteins

Protein	No. of peptides <sup>a</sup>	% Of subjects <sup>b</sup>	% Of peptides <sup>c</sup>
p17	17	60.0	94.1
p24	31	86.0	100
p15	18	50.7	94.4
Nef	27	91.3	96.3
Rev	15	32.7	93.3
Tat	12	28.7	83.3
Vpu	9	11.3	77.8
Pro	20	25.3	65.0
RT	76	72.7	90.8
Int	37	76.0	86.5
Vpr	11	37.3	100
gp120	67	65.3	83.6
gp41	46	57.3	78.3
Vif	24	40.0	100
Total	410	100	88.3

<sup>a</sup> Number of overlapping peptides spanning the protein sequence.<sup>b</sup> Percentage of subjects targeting each of the proteins.<sup>c</sup> Percentage of peptides recognized within each protein by at least one individual.

counted as two epitopic regions. All cases of responses to four, six, or eight adjacent peptides were considered multiple pairs and counted as two, three, or four epitopic regions, respectively. All cases where five and seven adjacent peptides were targeted were considered a triplet and one or two pairs and counted as three and four epitopic regions, respectively. The data set contained 1,638 responses to single peptides, 691 cases of responses to two adjacent peptides, 112 cases of responses to three adjacent peptides, 36 cases of responses to four adjacent peptides, 15 cases of responses to five adjacent peptides, 3 cases of responses to six adjacent peptides, 2 cases of responses to seven adjacent peptides, and 1 case of responses to eight adjacent peptides.

The total magnitude of responses was determined as the sum of the responses to all peptides and epitopic regions. The average magnitude was obtained by dividing the total magnitude by the number of targeted peptides.

Overall, the number of responses expressed as single peptides was highly significantly associated with the number of epitopic regions ( $P < 0.0001$ ;  $r = 0.9903$ ; Spearman rank correlation). Statistical analysis of correlations between the breadth or magnitude of responses and the viral load, as well as CD4 counts, yielded mostly identical  $P$  values when both counting approaches were applied. Therefore, data are presented based on counting the actual number of peptides that were recognized.

**Entropy, frequency of "forbidden" amino acids, and cleavage scores.** As a measure of the variability of each of the peptides, the average entropy scores for all 410 OLPs were determined. The average entropy score for each peptide was calculated from Shannon entropy for each residue of B clade protein alignments, using only complete proteins and only a single protein from multisequence sets derived from a single patient (36, 47). The fraction of forbidden residues was determined based on the number of G, P, E, D, Q, N, T, S, and C residues divided by the number of residues in the peptide. These amino acids rarely serve as the C-terminal anchors of optimally defined CTL epitopes (15). Proteasomal-cleavage prediction scores were determined based on a cleavage prediction algorithm (34) that has been shown to be highly predictive of C termini in HIV-1 epitopes (55). A cutoff score of  $\geq 0.7$  (based on comparisons to known epitopes in the database) (55) was used, and the fraction of amino acid residues that had a score above the threshold was determined. Nine amino acids on either side of the residue whose cleavage potential was being estimated were used to provide the context for the neural-net prediction (55). Only cleavage scores in amino acids in the C-terminal half of peptides were considered, starting at the eighth position in each peptide, thus avoiding high cleavage scores near the N terminus of a peptide, as high cleavage scores embedded within an epitope might have a counterbalancing effect of diminishing the number of intact epitopes processed for presentation. Stepwise multiple-regression analysis to model the contributions of these variables to predicting the frequency of recognition for each peptide was carried out using Splus version 6.1.2 (Insightful Corp.).

**Statistical analysis.** Statistical analysis was done using GraphPad Prism Macintosh version 3.0 and Splus version 6.1.2. The results are generally presented as median values with ranges. Tests included the nonparametric Mann-

Whitney test (two-tailed) for comparison between two ethnicities and treatment groups, the Kruskal-Wallis test for comparison among all four ethnicities, Spearman's nonparametric rank test (two-tailed) for correlations, and Fisher's exact test for the analysis of the HLA allele distributions. Correlations were considered statistically significant when  $P$  was  $< 0.05$ , unless multiple tests were performed, in which case Bonferroni correction was applied to correct for multiple subgroup analyses. Viral-load values below the limit of detection of 50 RNA copies/ml were assigned a value of 49 for statistical analyses; for values with a detection limit of 400 copies/ml, a value of 399 was used. Viral loads of  $> 750,000$  copies/ml were assigned a value of 750,001.

All data have been made available to the public and can be accessed at <http://www.hiv.lanl.gov/content/immunology/hlatem/index.html>.

## RESULTS

**Variable HLA allele frequencies in different ethnicities.** The presentation of antigenic peptides to CTL is largely dependent on the abilities of processed peptides to bind to a restricting HLA class I molecule. HLA allele frequencies vary among diverse ethnic populations, and this variance may have a direct influence on the distribution, breadth, and strength of virus-specific CTL responses. To address these issues, the MHC class I allele distribution was determined for a total of 275 HIV-infected and uninfected subjects from four different ethnicities: 73 African-Americans, 75 Caucasians, 47 Hispanics, and 80 subjects from the West Indies (Barbados and Haiti). HLA-A11 and HLA-A25 distributions differed significantly among ethnicities after correcting for multiple subgroup tests ( $P = 0.0001$  and  $P = 0.001$ , respectively) (Fig. 1). For seven additional HLA alleles (A03, A29, B27, B40, B57, Cw01, and Cw08), a trend toward different distributions among ethnicities was found ( $P < 0.05$ ).

**Broad distribution of HIV-specific CD8<sup>+</sup>-T-cell responses across all HIV proteins.** Previous studies have reported comprehensive mapping of CD8<sup>+</sup>-T-cell responses (1, 7, 41), but the cohort sizes have been limited and responses have not been linked to multiple ethnicities. To investigate which regions of the HIV protein sequences were targeted by the cellular immune response, 150 subjects of African-American, Caucasian, Hispanic, or West Indian descent were studied comprehensively for HIV-specific CD8<sup>+</sup>-T-cell activity. To this end, 410 OLPs based on a recent clade B consensus sequence (<http://hiv-web.lanl.gov/content/hiv-db/PEPTGEN/2001.html>) were used in a matrix-based IFN- $\gamma$  ELISpot assay as described previously (1).

Comprehensive screening for CD8<sup>+</sup>-T-cell responses identified an unprecedented breadth in the distribution of recognized peptides. Overall, the studied subjects responded to a median of 19 peptides, ranging from 1 to 67 recognized peptides. Gag, Nef, and integrase were the most frequently targeted proteins (Table 2). Overall, 88% (362 of 410) of all peptides spread over all proteins served as targets for CD8<sup>+</sup> T cells. The highest density of targeted peptides was seen for Gag p24, Vpr, and Vif, for which 100% of peptides were recognized by at least one person. The least broadly targeted protein was protease, with 65% of peptides spanning this region being targeted (Table 2).

**Identification of several highly targeted regions across diverse ethnicities.** Viral-sequence variability is considered a major roadblock in HIV vaccine development. However, vaccine design may also need to consider differences in HLA allele distribution among different ethnic populations, which may

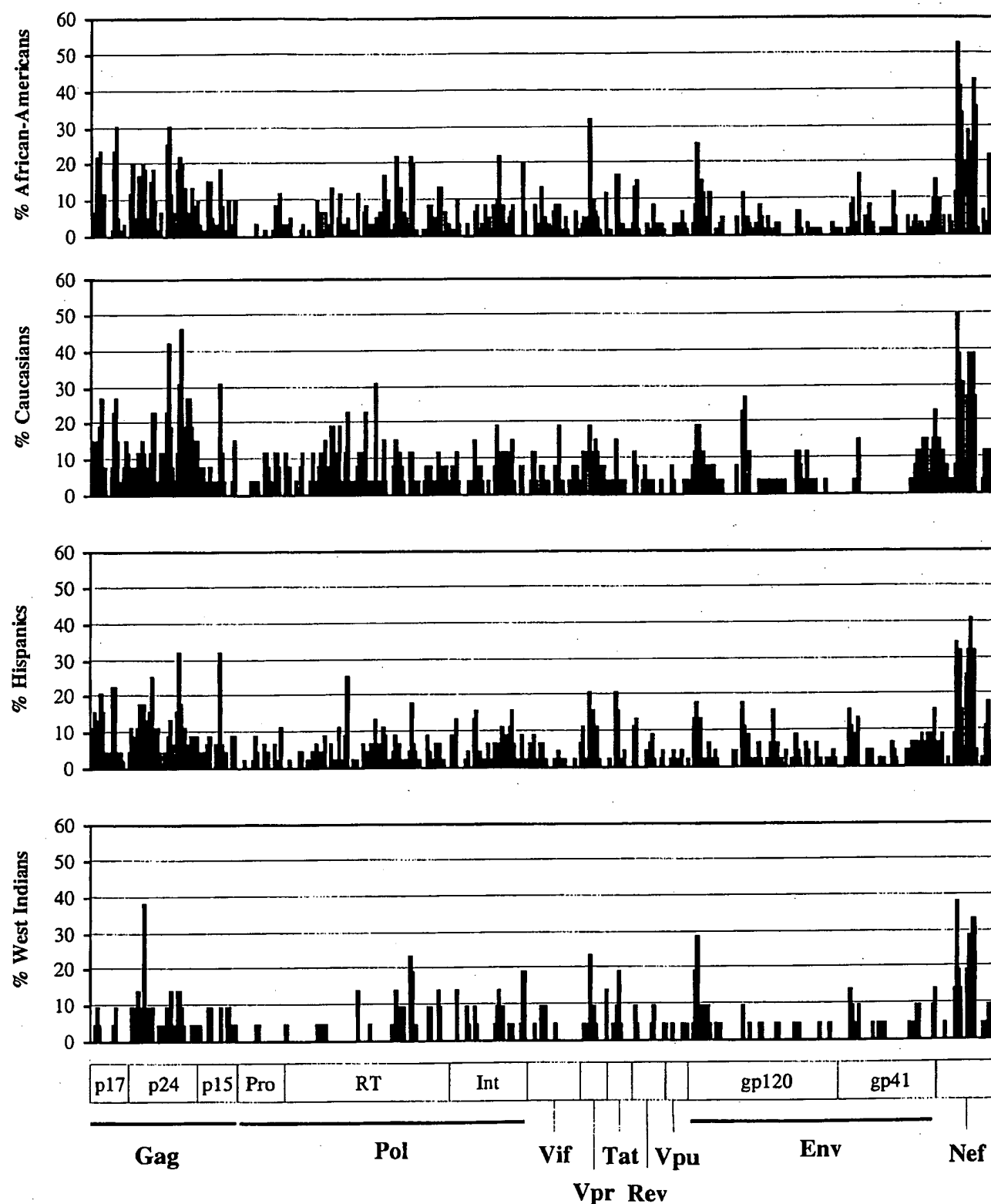


FIG. 2. CTL responses in four ethnic groups are distributed over the entire HIV genome and cluster in immunodominant regions. The total HIV-specific CD8<sup>+</sup>-T-cell activities for 59 African-Americans, 26 Caucasians, 44 Hispanics, and 21 West Indians were tested using an OLP set of 410 peptides. The frequencies of recognition for each single peptide in the four ethnicities are shown. The horizontal box at the bottom indicates the HIV proteins spanned by the OLPs.

TABLE 3. Peptides differentially targeted across ethnic groups

OLP	Protein	Sequence	Overall frequency of recognition	HLA <sup>a</sup>	Frequency of recognition per ethnicity <sup>b</sup>				P value
					AA	C	H	WI	
6	p17	ASRELERFAVNPGLL	13.3	A11	6.8	26.9	20.5	0.0	0.0068
17	p17/p24	AADTGNSSQVSONYPIV	2.7	A11, A25	0.0	15.4	0.0	0.0	0.001
35	p24	WMTNNPPIPVGEIYKRWI	16.0	B08, B53	25.4	23.1	2.3	9.5	0.0035
36	p24	PVGEIYKRWIILGLNKIV	22.0	B08, B27	30.5	42.3	9.1	4.8	0.001
194	RT	KIEELRQHLLRWGFTTPDK	13.3	B44	5.1	23.1	25.0	0.0	0.0018
207	RT	ELAENREILKEPVHGVYY	11.3	A2	5.1	30.8	13.6	0.0	0.0028

<sup>a</sup> HLA allele most commonly present among individuals with responses to OLP.<sup>b</sup> AA, African-American; C, Caucasian; H, Hispanic; WI, West Indian.

lead to the preferential targeting of some peptides in one ethnicity but not another (26). To address this point and to identify immunodominant regions that are equally or differentially targeted among different ethnicities, the distribution of responses was analyzed based on the ethnic descent of the tested subjects (Fig. 2). This analysis revealed a conserved pattern of responses for African-Americans, Caucasians, Hispanics, and West Indians, with all proteins targeted by all ethnicities and with particularly strong responses to HIV Gag and Nef. In all four ethnic cohorts, Nef had the most frequent responses, which clustered within two peaks around peptides 76 to 79 and 81 to 85.

Despite these strong similarities in the overall distribution of responses among the various ethnicities, six peptides in Gag-p17, Gag-p24, and RT were identified that were differentially targeted among groups ( $P < 0.01$ ) (Table 3). For each of these peptides, an association between the frequency of recognition in a given ethnic group and a potentially restricting HLA allele could be determined (Table 3).

**Identification of immunodominant regions for vaccine design.** While ethnicity-specific targeting of some peptides may occur, many peptides were frequently targeted by a substantial proportion of individuals across all ethnicities.

Peptides that were recognized by >15% of the total population are listed in Table 4. Nef and Gag stand out as containing the most frequently targeted peptides, particularly peptide Nef-76, which was recognized most frequently in African-Americans and West Indians and was also the second and fourth most frequently recognized peptide in Hispanics and Caucasians, respectively. Importantly, some of the 33 peptides shown in Table 4 also scored high when their intraindividual immunodominances were analyzed (data not shown). For instance, the Nef-76 epitope was recognized by almost 50% of the entire cohort, and it also had the strongest response for 15% of the individuals who targeted it.

In order to achieve population coverage regardless of HLA type, we combined heavily targeted peptides in immunodominant regions in an attempt to identify specific regions that are simultaneously targeted by a majority of individuals in each ethnicity. Table 5 indicates that as few as three such regions, two consecutive stretches in Nef and one shorter region in Gag-p24, are sufficient to achieve cumulative recognition by >90% of individuals in each ethnicity.

Although these three regions were sufficient to elicit responses in the vast majority of our study cohort, an effective vaccine may need to induce broader responses to overcome sequence heterogeneity and CTL escape (3). We therefore

extended our analyses to a total of 10 regions, each recognized by at least 20% of individuals and elongated by nine amino acids at the C terminus to facilitate proteasomal processing (34, 55) and which therefore may represent potential vaccine candidates for the vaccination of heterogeneous ethnic populations in regions where clade B is endemic (Table 6).

**Treatment success reduces the breadth and magnitude of responses.** The notion that treated subjects show narrower and weaker CD8<sup>+</sup>-T-cell responses (4, 29, 30, 32) was not confirmed in this cohort when the treated group was analyzed as a

TABLE 4. Frequently targeted and immunodominant regions in different ethnicities

OLP	Protein	Sequence	Total	% Recognition <sup>a</sup>			
				AA	C	H	WI
3	p17	EKIRLRPGGKKYKCLKHI	16.7	22.0	15.4	15.9	4.8
5	p17	KHIVWASRELERFAV	16.7	23.7	19.2	9.1	4.8
11	p17	TGSEELRSYNTVATLY	22.0	23.7	23.1	22.7	4.8
12	p17	SLYNTVATLYCVHQRIEV	25.3	30.5	26.9	22.7	9.5
23	p24	AFSPEVPMFSALSEGA	15.3	16.9	11.5	18.2	9.5
24	p24	PMFSALSEGATPQDLNTM	16.0	20.3	15.4	13.6	4.8
25	p24	GATPQDLNTMLNTVGGH	20.0	18.6	11.5	18.2	38.1
29	p24	AAEWDRLHPVHAGPIA	20.0	18.6	23.1	25.0	9.5
35	p24	WMTNNPPIPVGEIYKRWI	16.0	25.4	23.1	2.3	9.5
36	p24	PVGEIYKRWIILGLNKIV	22.0	30.5	42.3	9.1	4.8
40	p24	GPKEPFRDYVDRFYKTLR	16.7	18.6	11.5	15.9	14.3
41	p24	YVDRFYKTLRAEQASQEV	25.3	22.0	30.8	31.8	14.3
42	p24	LRAEQASQEVKNWMTETL	22.7	20.3	46.2	18.2	9.5
59	p15	ROANFLGKIWPSHKGR	23.3	18.6	30.8	31.8	9.5
76 <sup>b</sup>	Nef	EVGFVRPQVPLRPMTYK	44.7	52.5	50.0	34.1	38.1
77	Nef	QVPLRPMTYKAAVDLSHF	34.0	40.7	38.5	29.5	19.0
78	Nef	YKAAVDLSHFLKEKGGL	30.0	33.9	30.8	31.8	14.3
79	Nef	SHFLKEKGGLGLIYSQK	18.0	20.3	30.8	15.9	0.0
81	Nef	QKRQDILDVWVYHTQGYF	26.0	28.8	26.9	25.0	19.0
82	Nef	LWVYHTQGYFPDWQNY	30.0	25.4	38.5	31.8	28.6
83	Nef	QGYFPDWQNYTPGPGIRY	30.7	23.7	38.5	40.9	19.0
84	Nef	NYTPGPGIRYPLTFGWCF	37.3	42.4	38.5	31.8	33.3
85	Nef	RYPLTFGWCFKLVPV	32.0	35.6	26.9	31.8	33.3
91	Nef	PEKEVLVWKFDLSRLAFHH	15.3	22.0	11.5	18.2	9.5
115	Tat	FHCQVCFTTKGLGISYGR	16.7	16.9	15.4	20.5	9.5
216	RT	QKIATESIVWGKTPKFK	16.0	22.0	15.4	9.1	14.3
223	RT	QLEKEPIVGAETFYVDGA	15.3	22.0	11.5	4.5	23.8
224	RT	GAETFYVDGAANRETKL	18.0	20.3	11.5	18.2	19.0
275	RT	KVVPRRKAKIIRDYQKQM	15.3	20.3	7.7	9.1	19.0
281	Vpr	ELKNEAVRHFPRIWLHSL	25.3	32.2	19.2	20.5	23.8
294	gp120	TVYYGVVPVWKEATTTLF	22.7	25.4	19.2	18.2	28.6
314	gp120	YRLISCNTSVITQACPKV	15.3	11.9	23.1	18.2	9.5
401	gp41	HIPRRIRQGLERALL	16.7	15.3	23.1	15.9	14.3

<sup>a</sup> Values are given as percent recognition of peptide by overall (total) and ethnic cohorts. AA, African-American; C, Caucasian; H, Hispanic; WI, West Indian.<sup>b</sup> Peptide 76 is the most frequently targeted peptide across ethnicities.

TABLE 5. Three frequently targeted regions are recognized by &gt;90% of subjects across ethnic groups

Region	Sequence	Length <sup>a</sup>	% Recognition <sup>b</sup>				
			Total	AA	C	H	WI
I	EVGFVPRPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIYSQK	41	92.7	91.5	100	90.9	90.5
II	QKRQDILDWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVVPV	45					
III	PRTLNAWVKVVEEKAFSPEVIMFSALEGA	31					

<sup>a</sup> Number of amino acids.<sup>b</sup> Cumulative percentage of subjects recognizing regions I to III. AA, African-American; C, Caucasian; H, Hispanic; WI, West Indian.

whole (Fig. 3). Only when the treated subjects were stratified into a treatment-suppressed group (including subjects with a plasma viral load of  $\leq 200$  RNA copies/ml) and those who were treated but had viral loads of  $> 200$  RNA copies/ml did the previously described decline in the breadth and magnitude of CD8<sup>+</sup>-T-cell responses become apparent. While no difference between the breadths of responses in untreated and treated but unsuppressed subjects was observed, suppressive treatment led to a significant reduction in the number of targeted peptides ( $P = 0.0007$ ). Similarly, the magnitude of the total CD8<sup>+</sup>-T-cell response was also significantly ( $P < 0.0001$ ) reduced in the latter group compared to untreated subjects and individuals for whom treatment failed.

**CD4 count, but not viral load, correlates with breadth of CD8<sup>+</sup>-T-cell responses.** The observation that T-cell activity is diminished during successful therapy with very low or undetectable viral replication suggests that CTL responses and the viral load could be positively correlated. However, a number of reports have been published over the last few years that provide conflicting data on the correlation between the CTL activity, viral load, and CD4 counts (1, 7, 13, 14, 18, 19, 22, 25, 42, 45, 53). We therefore took advantage of the large number of untreated individuals in our cohort with a wide distribution in viral loads and CD4 counts to reassess these associations.

While no correlation was found between the viral load and either the breadth or magnitude of the responses (Fig. 4A and C), the CD4 count was associated with the number of targeted peptides ( $P = 0.0032$ ;  $r = 0.38$ ) (Fig. 4B) and showed a trend toward correlation with the total magnitude of responses ( $P = 0.0666$ ) (Fig. 4D).

The frequency of peptide recognition is influenced by the relative conservation of the peptide, immunoproteasomal cleavage sites, and C-terminal forbidden amino acids. Besides HLA

allele frequencies, factors such as viral sequence heterogeneity and processing preferences may contribute to the clustering of CTL responses. We took advantage of the extensive data set generated in this study to explore the factors that may determine peptide recognition by analyzing sequence heterogeneity, the distribution of immunoproteasomal cleavage sites, and amino acid residues known to impair epitope binding to class I molecules when present at the C terminus.

Sequence variability was addressed by creating a variability score based on the average Shannon entropy for each position in the peptide, using the B clade reference set from the HIV database at Los Alamos for our reference alignment. Figure 5 demonstrates that on a genomewide level, conserved peptides displaying low entropy scores were more frequently targeted than variable peptides.

Enrichment for increased predicted immunoproteasomal scores (34) in the C-terminal halves of peptides would be expected to increase the generation of optimal peptides with a variety of C termini which would bind to different HLA alleles. On the other hand, an increased frequency of amino acid residues rarely seen at the C termini of optimally defined CTL epitopes (forbidden amino acids—G, P, E, D, Q, N, T, S, and C) (9, 15, 21) would be expected to limit the number of optimal epitopes that could be generated from a longer peptide sequence. The data presented in Fig. 6A to C show that the cleavage preference was directly correlated with the frequency of recognition of a given peptide ( $P = 3 \times 10^{-4}$ ;  $r = 0.18$ ), while the content of forbidden amino acids and the variability of a peptide were inversely correlated with the frequency of recognition ( $P = 1.2 \times 10^{-5}$ ,  $r = -0.21$ , and  $P = 1.5 \times 10^{-7}$ ,  $r = -0.25$ , respectively). The three described variables did not correlate with each other, although there was a trend toward an inverse correlation between the cleavage score and the

TABLE 6. Highly reactive regions to be included in vaccine design<sup>a</sup>

OLP	Protein	Sequence	Length <sup>b</sup>
11–12	Gag-p17	TGSEELRSLYNTVATLYCVHQRIEVKDTKEALEK	34
22–25	Gag-p24	WVKVVEEKAFSPEVIMFSALESEGATPQDLNTMLNTVGGHQAAMQMLKE	49
28–30	Gag-p24	LKETINEEAAEWDRHPVHAGPIAPGOMREPRGSDIAGTTS	41
35–37	Gag-p24	WMTNNPIPVGIEYKRWIILGLNKIVRMYSPTSILDIRQGPKE	43
40–42	Gag-p24	GPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDC	43
59	Gag-p15	TRRQANFLGKIWP SHKGRPGNFLOSRP	27
281	Vpr	ELKNEAVRHFPRIWLHSLGQHIYETYG	27
293–295	gp120	AAEQWLWTVVYGVVPVWKEATTTLCASDACAYDTEVHNVWA	41
75–79	Nef	WLEAQEEEEVGFPVRPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIYSQKRQDILDWV	58
81–85	Nef	QKRQDILDWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPEPEKVEEAN	25

<sup>a</sup> Peptides were included when recognized by at least 20% of the overall population, and adjacent peptides were added when targeted by >10% of the overall cohort.<sup>b</sup> Number of amino acids.

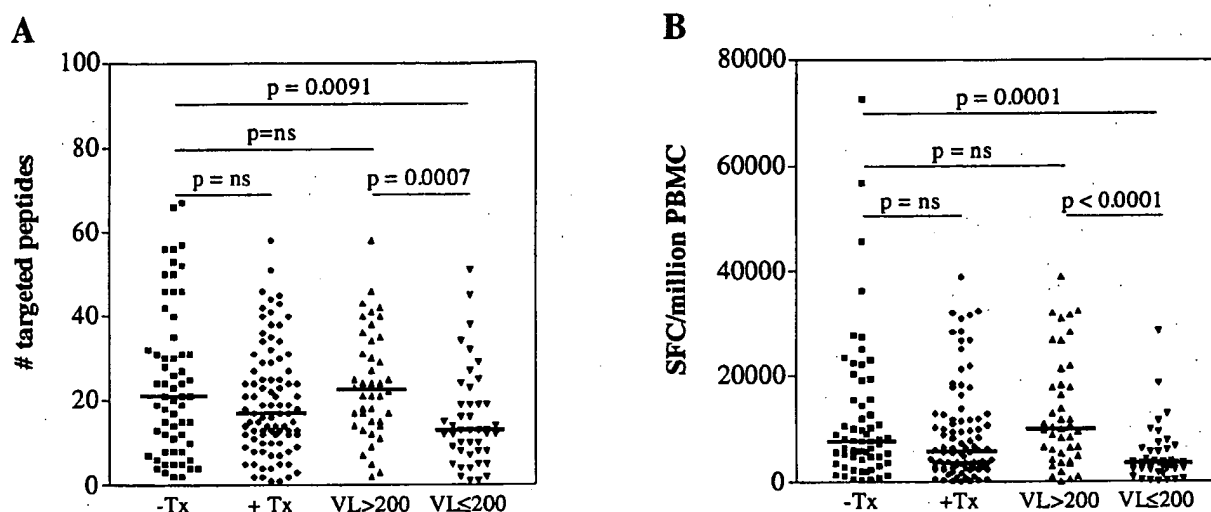


FIG. 3. Effective antiretroviral treatment reduces CTL activity. ELISpot responses are shown for 61 untreated (-Tx; squares) and 89 treated (+Tx; diamonds) subjects. The latter were further divided into 42 non-treatment-suppressed (viral load [VL] > 200; triangles) and 44 treatment-suppressed (VL ≤ 200; inverted triangles) individuals. Shown are the breadth of responses (number of targeted peptides) (A) and the total magnitude (number of SFC per million PBMCs) (B) for the different groups. The thick lines represent the median values for each group. *P* values were determined by Mann-Whitney tests.

entropy ( $P = 0.03$ ;  $r = -0.10$ ), suggesting that greater variability is associated with less efficient processing. Importantly, while each factor is independently related to the T-cell response—the effect of the entropy was more pronounced than those of the forbidden residues and the cleavage scores—the presence of two or more of these factors is cumulatively predictive for the frequency of recognition (Fig. 6D). Thus, the data indicate that peptides displaying low entropy, a low fraction of forbidden amino acids, and a high cleavage score are most frequently recognized in these *in vitro* studies, and antigens with these characteristics may represent potent vaccine candidates.

## DISCUSSION

Since the first description of HIV-specific CTL, numerous studies have supported the important role of this arm of the cellular immune system, and a considerable number of vaccine development efforts are directed toward the design of CTL-based vaccination strategies (10, 39, 54). Not only do global and local viral sequence diversities represent a major hurdle for vaccine development, but host genetics, especially HLA distribution, may also need to be taken into account when designing vaccine candidates to elicit strong CTL responses. In order to identify immunodominant regions in HIV that are targeted largely independently of HLA diversity, we assessed total HIV-specific CD8<sup>+</sup>-T-cell responses in a cohort of 150 HIV clade B-infected individuals of African-American, Caucasian, Hispanic, and West Indian descent, thus including the ethnicities in which the HIV epidemic progresses fastest in the United States (31). Total CD8<sup>+</sup>-T-cell IFN- $\gamma$  ELISpot responses were found to correlate with individuals' CD4 counts, but not viral loads, although CTL responses were higher in untreated individuals and individuals who obtained treatment but were unable to control viral replication. The single-peptide-based screening of this genetically diverse cohort allowed the identification of 10 regions that were heavily targeted in the

overall cohort and thus represent responses that are not restricted to a certain ethnicity. The present study also identifies three mechanisms which, independently from each other, predict the likelihood of single peptides being recognized in these *in vitro* studies. These factors—peptide entropy, the presence of immunoproteasomal cut sites, and the absence of amino acids that impair binding of optimal epitopes to the restricting class I molecules—may thus be employed to identify immunodominant regions that could be included in vaccine design, not only for HIV, but also for other pathogens.

In order to assess the extent of ethnicity-specific responses that may be present in one ethnic group but absent in others and to investigate whether these responses are due to HLA diversity and potential differences in the HLA class I allele-specific binding motifs, frequently targeted responses were compared among ethnicities. Only a few HLA alleles were found to differ significantly in their distributions among the ethnicities studied, most prominently HLA A11 and HLA A25, and the analysis of ethnicity-based peptide targeting revealed only six peptides with significantly altered frequencies of recognition. In each case, specific HLA class I alleles could be implicated as likely mediating the observed effect, although their distributions among ethnicities did not reach statistical significance in most cases. It is possible that peptides may carry multiple overlapping epitopes that can be presented by different HLA alleles, and this redundancy may overshadow specific HLA differences in population frequencies.

Furthermore, the cohort was also analyzed for the presence of HLA alleles associated with slow or fast disease progression (reviewed in reference 52). While ethnic groups with a high representation of these HLA alleles did not show significant differences in the viral loads of untreated individuals, our data confirmed the association of HLA-B57 with low plasma viral loads when the analysis was performed across ethnicities.

This study, the first to determine the total HIV-specific



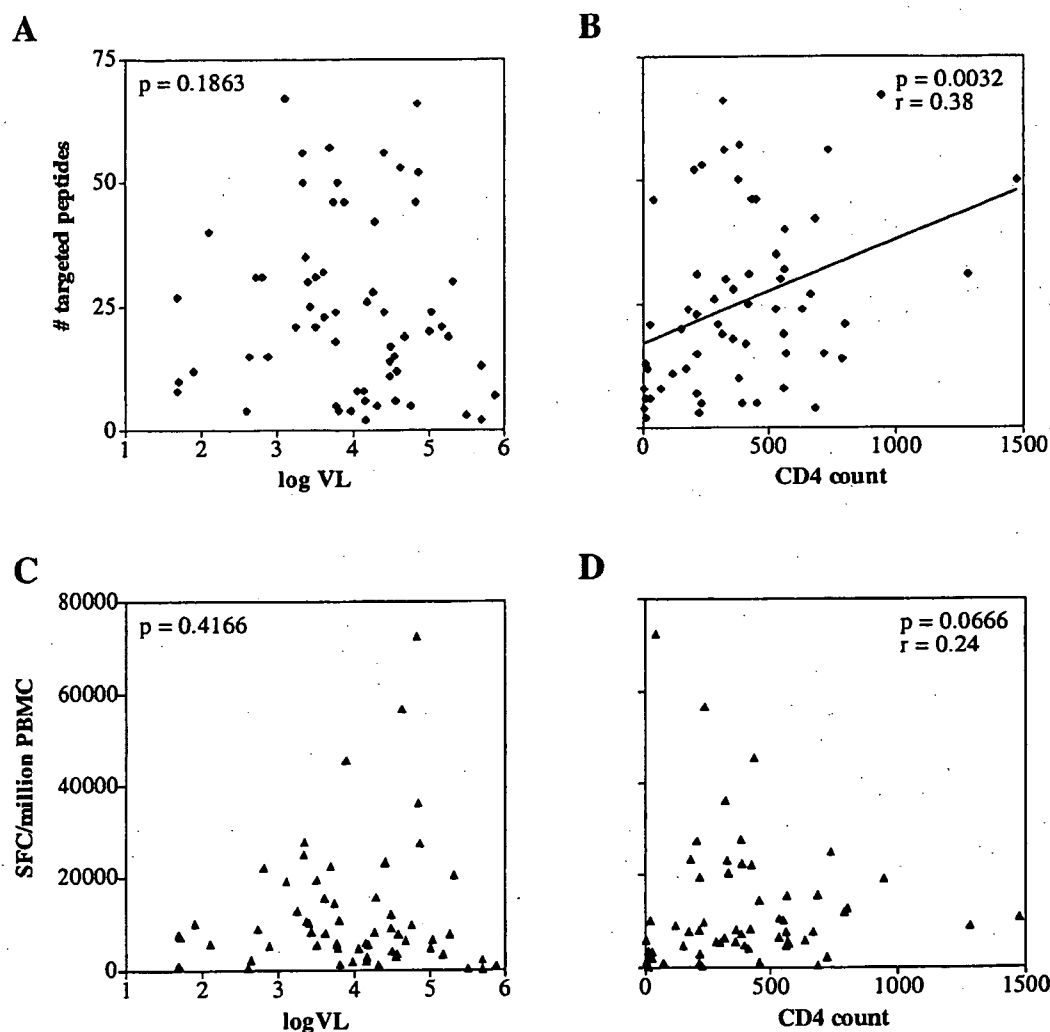


FIG. 4. CTL activity correlates with CD4 counts but not viral loads in untreated subjects. Association between the number of targeted peptides and the viral load (VL) (A) or CD4 count (B) and association between the total magnitude of the CD8<sup>+</sup>-T-cell response and viral load (C) or CD4 count (D) are shown for 61 untreated subjects. *P* values were determined by Spearman rank tests. Correlation in panel B remains significant when subjects with CD4 counts of >1,000 cells/ $\mu$ l are excluded (*P* = 0.0085). CD4 counts were available for 59 subjects.

CD8<sup>+</sup>-T-cell response in a large and predominantly non-Caucasian cohort, revealed an unprecedented breadth in the distribution of responses. All ethnicities were shown to mount CTL responses against every HIV protein, and <12% of the peptides were not targeted by any individual. This supports the usefulness of a consensus sequence-based approach, in which the test peptides on average resemble the individual's autologous virus more closely than peptide sets based on specific viral-isolate sequences (24, 55).

The present study included individuals with and without current antiretroviral treatment and with a wide range of viral loads and CD4 counts, allowing correlations of these parameters with immune responses in a large number of subjects. Surprisingly, when the breadths and magnitudes of responses in treated and untreated individuals were compared, no difference in CD8<sup>+</sup>-T-cell activity was observed. Only when the treated group was further stratified into treatment-suppressed and non-treatment-suppressed individuals were significant differences found in the breadth and magnitude of the virus-

specific T-cell responses. These were significantly higher in both the untreated and the non-treatment-suppressed groups than in individuals controlling their viral loads to  $\leq 200$  RNA copies/ml, extending earlier reports demonstrating that CTL responses decline under effective treatment and may, at least partially, fall below the limit of detection of current assays (4, 29, 30, 32).

Although this finding would suggest that the antigen load is the driving force for the immune response, we were unable to establish a correlation between the plasma viral load and either the breadth or magnitude of HIV-specific CD8<sup>+</sup>-T-cell responses, in contrast to previously published reports from other laboratories, which have reported positive (7, 13, 42) or negative (14, 18, 22, 42, 45, 53) correlations between these parameters. However, many of these earlier studies were performed on smaller cohorts and were less comprehensive than the present study. This does not, however, rule out the possibility that in the large amount of data generated by these comprehensive screenings, relevant immune responses may be ob-

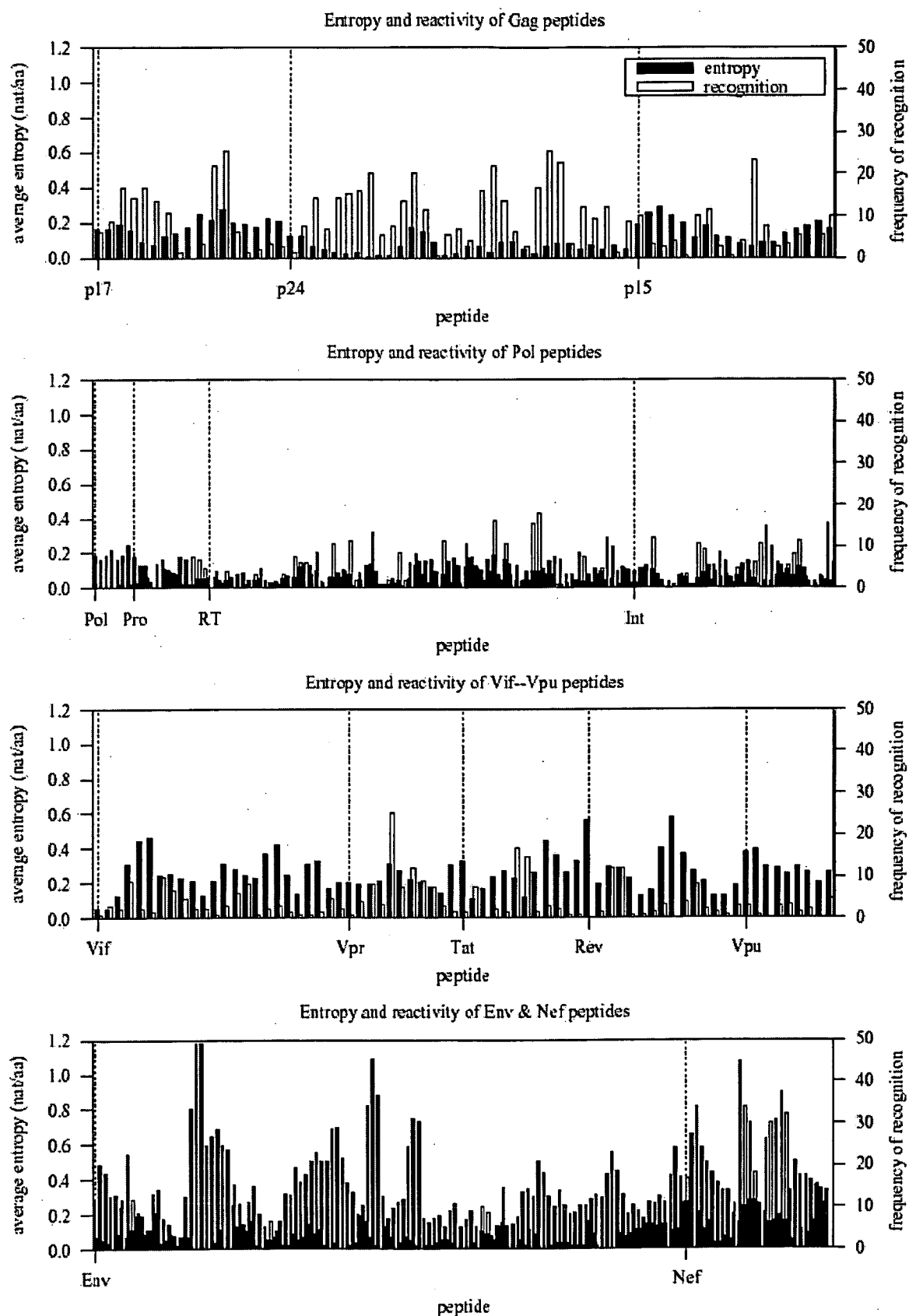


FIG. 5. The frequency of peptide recognition is inversely correlated with peptide entropy. The number of subjects recognizing each peptide and the peptide entropy for each of 410 OLPs were correlated ( $P < 0.0001$ ; Spearman rank test). Reactivity is shown by the open bars; entropy is shown by the solid bars. nat, natural unit; aa, amino acid.

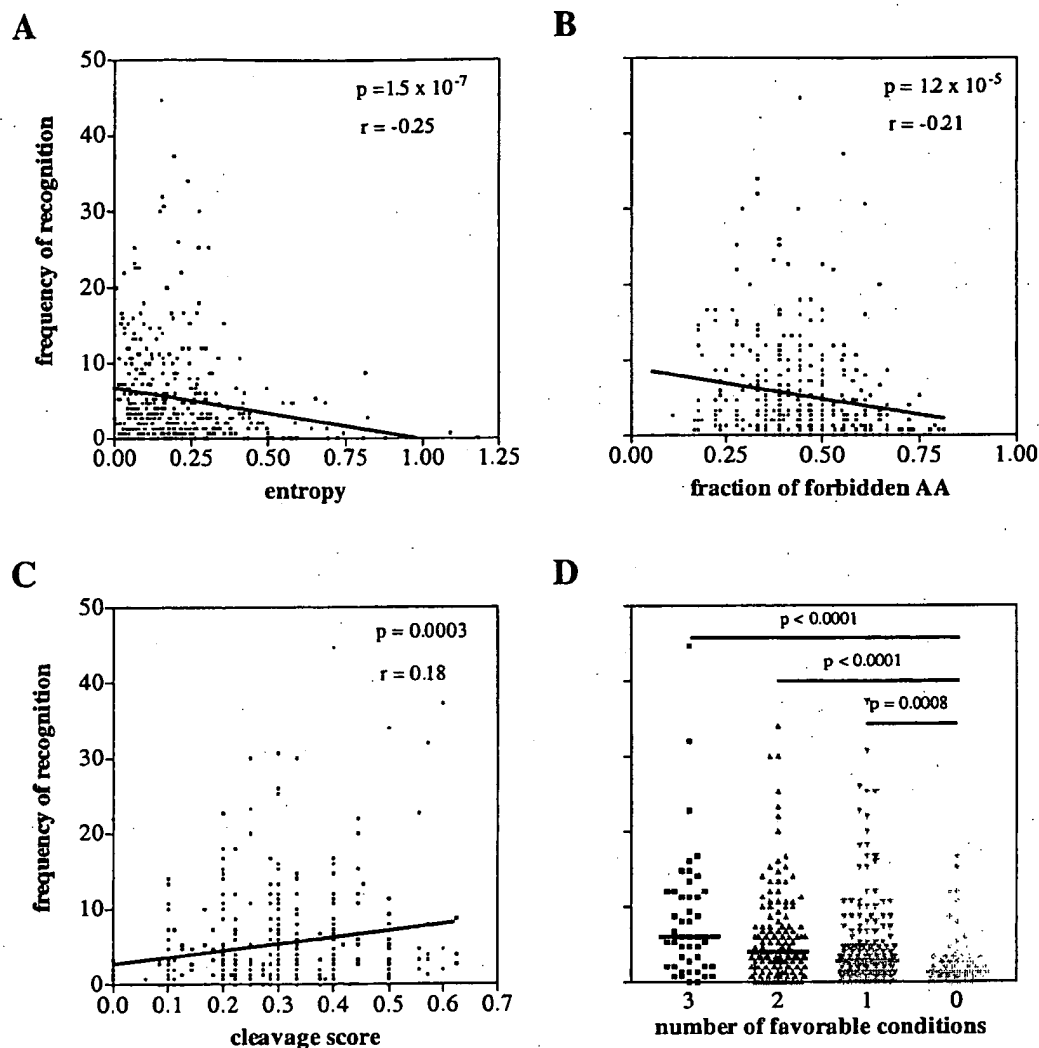


FIG. 6. (A to C) Influence of entropy (A), fraction of forbidden amino acids (AA) (B), and proteasomal cleavage scores (C) on the frequency of recognition of OLPs. Entropy and the fraction of forbidden amino acids correlate inversely ( $P < 0.0001$ ) and cleavage scores correlate positively ( $P = 0.0003$ ) with the frequency of recognition. (D) Favorable conditions for high recognition are defined as low entropy, few forbidden amino acids, and high cleavage scores; scores are divided into high and low categories at the median. Stepwise multiple-regression analyses indicated that each of the three variables contributed significantly to the model, that there were no second-order effects based on combinations of the variables, and that there was no benefit in using a quadratic model rather than a linear model to fit the data.

scored, and that some specificities may indeed correlate with viral loads. A further factor that should be considered is that our study was cross-sectional and the relative times from infection when the samples were obtained were not known. In addition, the possibility that responses to autologous virus were missed by using the reference strain of HIV employed here must be considered (2).

While the viral load was not associated with CD8<sup>+</sup>-T-cell activity, CD4 counts were found to positively correlate with the number of targeted peptides. Overall, there was also a trend toward a correlation between the magnitude of CD8<sup>+</sup>-T-cell responses and the CD4 count which was statistically significant when only the responses against the more conserved proteins, Gag and Pol, were considered, partly supporting the findings of Edwards et al. (22), who reported a positive correlation between Gag responses and CD4 counts. Thus, while antigen

presence above a threshold seems to be required to maintain a full CD8<sup>+</sup>-T-cell response, CD4 counts are, on a large-population level, a better predictor of the breadth and magnitude of HIV-specific CD8<sup>+</sup>-T-cell responses. However, the mechanisms behind this association remain unclear, as broad CTL responses could be crucial to maintaining CD4<sup>+</sup> T cells, or an elevated CD4 count could be necessary to enable the immune system to mount a broad CD8<sup>+</sup>-T-cell response. Further analysis will be necessary to understand the relationship between the CD4 count and CTL responses, especially studies addressing the antigen specificities of these CD4<sup>+</sup> T cells.

Overall, the present data demonstrate that almost the entire HIV protein sequence can be targeted by CD8<sup>+</sup> T cells. Thus, all HIV proteins are subject to antigen processing and presentation and could be included in HIV vaccines. However, despite this wide distribution of responses, the targeted peptides

seemed to cluster in highly reactive regions, especially within the HIV Gag and Nef proteins. This pattern was remarkably conserved among all four ethnic groups, allowing the identification of 10 reactive regions, each recognized by at least 20% of the overall cohort. These clusters were so consistently targeted between ethnicities that the combination of as few as three regions, each comprising 30 to 45 amino acids, was sufficient to achieve population coverage of >90%. As these regions represent potential vaccine candidates, it will be interesting to assess whether the same regions will be targeted in individuals of additional ethnicities or individuals infected with viruses other than clade B. Indeed, 6 of these 10 highly immunogenic clusters were also reported in the extensive studies of HIV clade C-infected individuals in Botswana performed by Novitsky et al. (41). These findings strongly suggest that factors other than host genetics may dictate which regions of the virus are immunogenic for HLA class I-restricted T cells.

There are several factors that could influence the observed CTL epitope clustering, among them sequence heterogeneity (9, 26, 38, 55). Yusim et al. used the complete set of optimal HIV-derived CTL epitopes included in the Los Alamos HIV Immunology Database to show that regions of HIV-1 proteins that were epitope rich (i) were conserved, (ii) had a high propensity to be cleaved by the immunoproteasome, and (iii) had a lower frequency of amino acids that do not serve as C-terminal anchor residues compared to regions of proteins that had a paucity of CTL epitopes (55). However, the study was based on tallying how many previously defined unique epitopes overlapped in a region and was unable to address how these factors related to the frequency of recognition. In contrast, in the present study, we were able to directly assess whether sequence entropy, the presence of proteasomal cleavage sites, or the frequency of forbidden residues correlated with the frequency at which a specific peptide was targeted; we applied a systematic approach to a genetically heterogeneous population, and our peptides completely spanned all HIV proteins for a comprehensive comparison. The three factors variability, proteasomal cleavage propensity, and low frequencies of forbidden amino acids were all found to be strong and independent predictors of the number of detectable CTL responses directed at a peptide. Importantly, since sequence entropy was not the only factor that predicted peptide reactivity, using autologous peptide test sequences might not completely eliminate epitope clustering, and autologous test sequences should still produce fewer reactive peptides in regions with selectively reduced proteasomal cleavage sites and HLA class I binding potential.

The relatively reduced number of reactive peptides in variable regions could be in part a consequence of these regions being less functionally constrained. Specific CTL escape mutations in variable regions may have a greater capacity to cumulatively diminish the overall potential for epitope processing and presentation at the population level (40, 55), as the fitness costs of escape may be less able to drive reversion to the susceptible form upon transmission to a new host, where specific immune pressure is lifted. This concept is supported by the finding that the predicted frequency of proteasomal cleavage sites showed a trend toward an inverse correlation with peptide entropy, suggesting that sequence variability may reflect viral adaptation and antigen-processing escape. These

three correlates of CTL immunogenicity combined may drive the observed epitope clustering, allowing the identification of interindividually highly immunodominant peptides with patterns that hold across different ethnicities.

Taken together, the results of the present study, performed on an extensive population with diverse genetic backgrounds, describe highly immunogenic peptides that may be incorporated into HIV vaccine candidates independently of host genetics. The study also provides mechanisms that are at least partially responsible for the in vitro immunodominance and likely high in vivo immunogenicity of these peptides. A combination of these regions or proteins, along with additional vaccine components that are able to induce strong T helper cell responses and potentially neutralizing antibodies, should contribute to the design of an effective HIV vaccine (48, 50).

#### ACKNOWLEDGMENTS

This work was funded by National Institutes of Health contract N01-AI-15422 (HLA typing and CTL epitope mapping to guide HIV vaccine development).

We gratefully acknowledge helpful discussion with and guidance by Patricia D'Souza for this project and thank Marcus Altfeld, Tonia Woodberry, and Todd Suscovich for their critical reviews of the manuscript. We thank Nancy David and James Theiler for helpful discussions regarding the strategies we used for statistical analysis.

The Los Alamos National Laboratory control number is la-ur-03-5892.

#### REFERENCES

1. Addo, M. M., X. G. Yu, A. Rathod, D. Cohen, R. L. Eldridge, D. Strick, M. N. Johnston, C. Corcoran, A. G. Wurcel, C. A. Fitzpatrick, M. E. Feeney, W. R. Rodriguez, N. Basgoz, R. Draenert, D. R. Stone, C. Brander, P. J. R. Goulder, E. S. Rosenberg, M. Altfeld, and B. D. Walker. 2003. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* 77:2081-2092.
2. Altfeld, M., M. M. Addo, R. Shankarappa, P. K. Lee, T. M. Allen, X. G. Yu, A. Rathod, J. D. Harlow, K. O'Sullivan, M. N. Johnston, P. J. R. Goulder, J. I. Mullins, E. S. Rosenberg, C. Brander, B. T. Korber, and B. D. Walker. 2003. Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. *J. Virol.* 77:7330-7340.
3. Altfeld, M., T. Allen, X. G. Yu, M. N. Johnston, D. Agrawal, B. T. Korber, D. C. Montefiori, D. H. O'Connor, B. T. Davis, P. K. Lee, E. L. Maier, J. D. Harlow, P. J. R. Goulder, C. Brander, E. S. Rosenberg, and B. D. Walker. 2002. HIV-1 superinfection despite broadly directed pre-existing CD8<sup>+</sup> T cell responses containing replication of the primary virus. *Nature* 420:434-439.
4. Altfeld, M., J. van Lunzen, N. Frahm, X. G. Yu, C. Schneider, R. L. Eldridge, M. E. Feeney, D. Meyer-Olson, H. J. Stellbrink, and B. D. Walker. 2002. Expansion of pre-existing, lymph node-localized CD8<sup>+</sup> T cells during supervised treatment interruptions in chronic HIV-1 infection. *J. Clin. Invest.* 109:837-843.
5. Amara, R. R., F. Villinger, J. D. Altman, S. L. Lydy, S. P. O'Neil, S. I. Staprans, D. C. Montefiori, Y. Xu, J. G. Herndon, L. S. Wyatt, M. A. Candido, N. L. Kozyr, P. L. Earl, J. M. Smith, H. L. Ma, B. D. Grimm, M. L. Hulsey, J. Miller, H. M. McClure, J. M. McNicholl, B. Moss, and H. L. Robinson. 2001. Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* 292:69-74.
6. Barouch, D. H., S. Santra, J. E. Schmitz, M. J. Kuroda, T. M. Fu, W. Wagner, M. Bilska, A. Craiu, X. X. Zheng, G. R. Krivulka, K. Beaudry, M. A. Lifton, C. E. Nickerson, W. L. Trigona, K. Punt, D. C. Freed, L. Guan, S. Dubey, D. Casimiro, A. Simon, M. E. Davies, M. Chastain, T. B. Strom, R. S. Gelman, D. C. Montefiori, M. G. Lewis, E. A. Emini, J. W. Shiver, and N. L. Letvin. 2000. Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* 290:486-492.
7. Betts, M. R., D. R. Ambrozak, D. C. Douek, S. Bonhoeffer, J. M. Brenchley, J. P. Casazza, R. A. Koup, and L. J. Picker. 2001. Analysis of total human immunodeficiency virus (HIV)-specific CD4<sup>+</sup> and CD8<sup>+</sup> T-cell responses: relationship to viral load in untreated HIV infection. *J. Virol.* 75:11983-11991.
8. Betts, M. R., J. P. Casazza, B. A. Patterson, S. Waldrop, W. Trigona, T. M.

- Fu, F. Kern, L. J. Picker, and R. A. Koup. 2000. Putative immunodominant human immunodeficiency virus-specific CD8<sup>+</sup>-T-cell responses cannot be predicted by major histocompatibility complex class I haplotype. *J. Virol.* 74:9144-9151.
9. Brander, C., and P. Goulder. 2001. The identification of optimal HIV derived CTL epitopes in diverse populations using HIV clade specific consensus, p. 1-1-1-20. In B. T. Korber, C. Brander, B. F. Haynes, R. Koup, C. Kuiken, J. P. Moore, B. D. Walker, D. I. Watkins, et al. (ed.), HIV molecular immunology database. Los Alamos National Laboratory, Los Alamos, N.Mex.
  10. Brander, C., and Y. Riviere. 2002. Early and late cytotoxic T lymphocyte responses in HIV infection. *AIDS* 16(Suppl. 4):S97-S103.
  11. Brander, C., and B. D. Walker. 1999. T lymphocyte responses in HIV-1 infection. Implications for vaccine development. *Curr. Opin. Immunol.* 11: 451-459.
  12. Bunce, M., C. M. O'Neill, M. C. Barnardo, P. Krausa, M. J. Browning, P. J. Morris, and K. I. Welsh. 1995. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 and DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46:355-367.
  13. Buseyne, F., J. Le Chenadec, B. Corre, F. Porrot, M. Burgard, C. Rouzioux, S. Blanche, M. J. Mayaux, and Y. Riviere. 2002. Inverse correlation between memory Gag-specific cytotoxic T lymphocytes and viral replication in human immunodeficiency virus-infected children. *J. Infect. Dis.* 186:1589-1596.
  14. Buseyne, F., D. Scott-Algara, F. Porrot, B. Corre, N. Bellal, M. Burgard, C. Rouzioux, S. Blanche, and Y. Riviere. 2002. Frequencies of ex vivo-activated human immunodeficiency virus type 1-specific gamma interferon-producing CD8<sup>+</sup> T cells in infected children correlate positively with plasma viral load. *J. Virol.* 76:12414-12422.
  15. Calef, C., R. Thakallapally, D. Lang, C. Brander, P. J. R. Goulder, O. Yang, and B. T. Korber. 2000. PeptGen: designing peptides for immunological studies and application to HIV consensus sequences, p. 1-63-1-100. In B. T. Korber, C. Brander, B. D. Walker, R. Koup, J. P. Moore, B. Haynes, G. Meyers, et al. (ed.), HIV molecular immunology database. Los Alamos National Laboratory, Los Alamos, N.Mex.
  16. Cao, J., J. McNevin, S. Holte, L. Fink, L. Corey, and J. McElrath. 2003. Comprehensive analysis of human immunodeficiency virus type 1 (HIV-1)-specific gamma interferon-secreting CD8<sup>+</sup> T cells in primary HIV-1 infection. *J. Virol.* 77:6867-6878.
  17. Cao, K., J. Hollenbach, X. Shi, W. Shi, M. Chopek, and M. A. Fernandez-Vina. 2001. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum. Immunol.* 62:1009-1030.
  18. Chouquet, C., B. Autran, E. Gomard, J. M. Bouley, V. Calvez, C. Katlama, D. Costagliola, and Y. Riviere. 2002. Correlation between breadth of memory HIV-specific cytotoxic T cells, viral load and disease progression in HIV infection. *AIDS* 16:2399-2407.
  19. Dalod, M., M. Dupuis, J. C. Deschemin, D. Sicard, D. Salmon, J. F. Del-fraissy, A. Venet, M. Sinet, and J. G. Guillet. 1999. Broad, intense anti-human immunodeficiency virus (HIV) ex vivo CD8<sup>+</sup> responses in HIV type 1-infected patients: comparison with anti-Epstein-Barr virus responses and changes during antiretroviral therapy. *J. Virol.* 73:7108-7116.
  20. Delwart, E. L., E. G. Shpaer, J. Louwagie, F. E. McCutchan, M. Grez, H. Rübsamen-Waigmann, and J. I. Mullins. 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. *Science* 262:1257-1261.
  21. Draenert, R., M. Altfeld, C. Brander, N. Basgoz, C. Corcoran, A. G. Wurcel, D. R. Stone, S. A. Kalamas, A. Trocha, M. M. Addo, P. J. R. Goulder, and B. D. Walker. 2003. Comparison of overlapping peptide sets for detection of antiviral CD8 and CD4 T cell responses. *J. Immunol. Methods* 275:19-29.
  22. Edwards, B. H., A. Bansal, S. Sabbaj, J. Bakari, M. J. Mulligan, and P. A. Goepfert. 2002. Magnitude of functional CD8<sup>+</sup>-T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *J. Virol.* 76:2298-2305.
  23. Feeney, M. E., K. A. Roosevelt, Y. Tang, K. J. Pfafferoth, K. McIntosh, S. K. Burchett, C. Mao, B. D. Walker, and P. J. R. Goulder. 2003. Comprehensive screening reveals strong and broadly directed human immunodeficiency virus type 1-specific CD8 responses in perinatally infected children. *J. Virol.* 77:7492-7501.
  24. Gaschen, B., J. Taylor, K. Yusim, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T. Bhattacharya, and B. Korber. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* 296:2354-2360.
  25. Gea-Banacloche, J. C., S. A. Migueles, L. Martino, W. L. Shupert, A. C. McNeil, M. S. Sabbaghian, L. Ehler, C. Prussin, R. Stevens, L. Lambert, J. Altman, C. W. Hallahan, J. C. de Quiros, and M. Connors. 2000. Maintenance of large numbers of virus-specific CD8<sup>+</sup> T cells in HIV-infected progressors and long-term nonprogressors. *J. Immunol.* 165:1082-1092.
  26. Goulder, P. J., C. Brander, K. Annamalai, N. Mngqundanisio, U. Govender, Y. Tang, S. He, K. E. Hartman, C. A. O'Callaghan, G. S. Ogg, M. A. Altfeld, E. S. Rosenberg, H. Cao, S. A. Kalamas, M. Hammond, M. Bunce, S. I. Pelton, S. A. Burchett, K. McIntosh, H. M. Coovadia, and B. D. Walker. 2000. Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and Caucasoid adults and children. *J. Virol.* 74:5679-5690.
  27. Goulder, P. J., C. Brander, Y. Tang, C. Tremblay, R. A. Colbert, M. M. Addo, E. S. Rosenberg, T. Nguyen, R. Allen, A. Trocha, M. Altfeld, S. He, M. Bunce, R. Funkhouser, S. I. Pelton, S. K. Burchett, K. McIntosh, B. T. Korber, and B. D. Walker. 2001. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* 412:334-338.
  28. Goulder, P. J., Y. Tang, C. Brander, M. R. Betts, M. Altfeld, K. Annamalai, A. Trocha, S. He, E. S. Rosenberg, G. Ogg, C. A. O'Callaghan, S. A. Kalamas, R. E. McKinney, Jr., K. Mayer, R. A. Koup, S. I. Pelton, S. K. Burchett, K. McIntosh, and B. D. Walker. 2000. Functionally inert HIV-specific cytotoxic T lymphocytes do not play a major role in chronically infected adults and children. *J. Exp. Med.* 192:1819-1832.
  29. Gray, C. M., J. Lawrence, J. M. Schapiro, J. D. Altman, M. A. Winters, M. Crompton, M. Loi, S. K. Kundu, M. Davis, and T. C. Merigan. 1999. Frequency of class I restricted anti-HIV-1 CD8 T cells in individuals receiving highly active antiretroviral therapy (HAART). *J. Immunol.* 162:1780-1788.
  30. Jin, X., G. Ogg, S. Bonhoeffer, J. Saffrit, M. Vesanen, D. Bauer, D. Chen, Y. Cao, M. A. Demoitie, L. Zhang, M. Markowitz, D. Nixon, A. McMichael, and D. D. Ho. 2000. An antigenic threshold for maintaining human immunodeficiency virus type 1-specific cytotoxic T lymphocytes. *Mol. Med.* 6:803-809.
  31. Joint United Nations Program on HIV/AIDS. 2002. AIDS epidemic update. Joint United Nations Program on HIV/AIDS, Geneva, Switzerland.
  32. Kalamas, S. A., P. J. Goulder, A. K. Shea, N. G. Jones, A. K. Trocha, G. S. Ogg, and B. D. Walker. 1999. Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. *J. Virol.* 73:6721-6728.
  33. Kaslow, R. A., M. Carrington, R. Apple, L. Park, A. Munoz, A. J. Saah, J. J. Goedert, C. Winkler, S. J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich, and D. L. Mann. 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat. Med.* 2:405-411.
  34. Kesmir, C., A. K. Nussbaum, H. Schild, V. Detours, and S. Brunak. 2002. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng.* 15:287-296.
  35. Klenerman, P., Y. Wu, and R. Phillips. 2002. HIV: current opinion in escapology. *Curr. Opin. Microbiol.* 5:408-413.
  36. Korber, B. T., K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, and S. M. Wolinsky. 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* 68:7467-7481.
  37. Koup, R. A., J. T. Saffrit, Y. Cao, C. A. Andrews, G. McLeod, W. Borkowsky, C. Farthing, and D. D. Ho. 1994. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J. Virol.* 68:4650-4655.
  38. Kuiken, C. L., B. Foley, E. Guzman, and B. T. Korber (ed.). 1999. The determinants of HIV-1 protein evolution. Johns Hopkins University Press, Baltimore, Md.
  39. McMichael, A., M. Mwau, and T. Hanke. 2002. Design and tests of an HIV vaccine. *Br. Med. Bull.* 62:87-98.
  40. Moore, C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296:1439-1443.
  41. Novitsky, V., H. Cao, N. Rybak, P. Gilbert, M. F. McLane, S. Gaolekwe, T. Peter, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee, and M. Essex. 2002. Magnitude and frequency of cytotoxic-T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. *J. Virol.* 76:10155-10168.
  42. Novitsky, V., P. Gilbert, T. Peter, M. F. McLane, S. Gaolekwe, N. Rybak, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee, and M. Essex. 2003. Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection. *J. Virol.* 77:882-890.
  43. O'Brien, S. J., X. Gao, and M. Carrington. 2001. HLA and AIDS: a cautionary tale. *Trends Mol. Med.* 7:379-381.
  44. O'Connor, D., T. Allen, and D. I. Watkins. 2001. Vaccination with CTL epitopes that escape: an alternative approach to HIV vaccine development? *Immunol. Lett.* 79:77-84.
  45. Ogg, G. S., X. Jin, S. Bonhoeffer, P. R. Dunbar, M. A. Nowak, S. Monard, J. P. Segal, Y. Cao, S. L. Rowland-Jones, V. Cerundolo, A. Hurley, M. Markowitz, D. D. Ho, D. F. Nixon, and A. J. McMichael. 1998. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* 279:2103-2106.
  46. Sabbaj, S., A. Bansal, G. D. Ritter, C. Perkins, B. H. Edwards, E. Gough, J. Tang, J. J. Szinger, B. Korber, C. M. Wilson, R. A. Kaslow, M. J. Mulligan, and P. A. Goepfert. 2003. Cross-reactive CD8<sup>+</sup> T cell epitopes identified in US adolescent minorities. *J. Acquir. Immune Defic. Syndr.* 33:426-438.
  47. Shannon, C. E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:623-656.

48. Shedlock, D. J., and H. Shen. 2003. Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science* 300:337–339.
49. Shiver, J. W., T. M. Fu, L. Chen, D. R. Casimiro, M. E. Davies, R. K. Evans, Z. Q. Zhang, A. J. Simon, W. L. Trigona, S. A. Dubey, L. Huang, V. A. Harris, R. S. Long, X. Liang, L. Handt, W. A. Schleif, L. Zhu, D. C. Freed, N. V. Persaud, L. Guan, K. S. Punt, A. Tang, M. Chen, K. A. Wilson, K. B. Collins, G. J. Heidecker, V. R. Fernandez, H. C. Perry, J. G. Joyce, K. M. Grimm, J. C. Cook, P. M. Keller, D. S. Kresock, H. Mach, R. D. Troutman, L. A. Isopi, D. M. Williams, Z. Xu, K. E. Bohannon, D. B. Volkin, D. C. Montefiori, A. Miura, G. R. Krivulka, M. A. Lifton, M. J. Kuroda, J. E. Schmitz, N. L. Letvin, M. J. Caulfield, A. J. Bett, R. Youil, D. C. Kaslow, and E. A. Emini. 2002. Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* 415:331–335.
50. Sun, J. C., and M. J. Bevan. 2003. Defective CD8 T cell memory following acute infection without CD4 T cell help. *Science* 300:339–342.
51. Trachtenberg, E., B. Korber, C. Sollars, T. B. Kepler, P. T. Hraber, E. Hayes, R. Funkhouser, M. Fugate, J. Theiler, Y. S. Hsu, K. Kunstman, S. Wu, J. Phair, H. Erlich, and S. Wolinsky. 2003. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* 9:928–935.
52. Trachtenberg, E. A., and H. A. Erlich. 2001. A review of the role of the human leukocyte antigen (HLA) system as a host immunogenic factor influencing HIV transmission and progression to AIDS, p. 1–43–1–60. *In* B. T. K. Korber, C. Brander, B. F. Haynes, J. P. Moore, R. A. Koup, C. Kuiken, B. D. Walker, and D. I. Watkins, et al. (ed.), *HIV molecular immunology 2001*. Los Alamos National Laboratory, Los Alamos, N.Mex.
53. van Baalen, C. A., O. Pontesilli, R. C. Huisman, A. M. Geretti, M. R. Klein, F. de Wolf, F. Miedema, R. A. Gruters, and A. D. Osterhaus. 1997. Human immunodeficiency virus type 1 Rev- and Tat-specific cytotoxic T lymphocyte frequencies inversely correlate with rapid progression to AIDS. *J. Gen. Virol.* 78:1913–1918.
54. Walker, B. D., S. Chakrabarti, B. Moss, T. J. Paradis, T. Flynn, A. G. Durno, R. S. Blumberg, J. C. Kaplan, M. S. Hirsch, and R. T. Schooley. 1987. HIV-specific cytotoxic T lymphocytes in seropositive individuals. *Nature* 328:345–348.
55. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigae, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.* 76:8757–8768.